Supporting Information for "Heat-stress increase under climate change twice as large in cities as in rural areas"

Hendrik Wouters[1], Koen De Ridder[2], Lien Poelmans[2], Patrick Willems[3], Johan Brouwers[4], Parisa Hosseinzadehtalaei[3], Hossein Tabari[3], Sam Vanden Broucke[1], Nicole P.M. van Lipzig[1], Matthias Demuzere[1,5]

Contents of this file

Text S1: The COSMO-CLM model. The COSMO model [*Steppeler et al.*, 2003] is a 3D atmospheric numerical model designed for operational and research applications in limited-area weather prediction at high resolution. It has been developed by the German Weather Service (DWD) and maintained by the Consortium for Small-Scale Modelling (`http://www.cosmo-model.org`). The model has a compressible non-hydrostatic core for atmospheric dynamics, and includes parametrizations for radiative transfer, cloud microphysics, subgrid-scale turbulence and convection [*Doms et al.*, 2011]. The ground heat- and water transport and the representation of vegetation and snow-cover are resolved by the Soil-Vegetation-Atmosphere Transfer (SVAT) module TERRA_ML [*Schulz et al.*, 2016; *Doms et al.*, 2011; *Grasselt*, 2008]. The land-atmosphere interactions are represented by the next generation TKE-based surface-layer transfer scheme [*Doms et al.*, 2011; *Buzzi*,

2008]. The regional climate model COSMO-CLM (COSMO model in CLimate Mode) is based on the COSMO model, and includes modifications allowing the application on time scales up to centuries [*Rockel et al.*, 2008]. These modifications comprise the introduction of time-dependent vegetation parameters and CO_2 concentration in the atmosphere. The COSMO-CLM model is developed and maintained by a vast amount of researchers of the COSMO consortium and the CLM-community (`http://www.clm-community.eu`). The model has been extensively used for climate downscaling applications on the continental and regional scale [*Thiery et al.*, 2016; *Vanden Broucke et al.*, 2015; *Akkermans et al.*, 2014; *Davin et al.*, 2014] down to the convection-permitting scale [*Grossman-Clarke et al.*, 2016; *Brisson et al.*, 2016a; *Prein et al.*, 2013]

In order to capture the urban-atmosphere interactions and associated atmospheric features including urban heat islands, the urban land-surface scheme TERRA_URB [*Wouters et al.*, 2016, 2015; *Demuzere et al.*, 2017] was introduced in the COSMO(-CLM) model. It provides an urban upgrade of its soil module TERRA_ML, the land-atmosphere interactions, and the land-surface parameters. As such, it represents the variability of ground heat and moisture transport, the turbulent transfer of momentum, heat and moisture, and the surface-atmosphere radiative exchanges in urban areas. The initial release of TERRA_URB features efficient implementations of the surface-layer stability functions accounting for the roughness sub-layer [*Wouters et al.*, 2012] and the impervious water-storage modelling based on a probability density function of water reservoirs [*Wouters et al.*, 2015]. Anthropogenic heat emissions are included as an additional heat source to the first above-ground model layer. Hereby, the magnitude takes into account latitude-dependent seasonal and diurnal distribution functions [*Flanner*, 2009] that are superim-

posed on the annual-mean anthropogenic heat flux. The latest version of TERRA_URB also implements the Semi-empirical Urban canopY dependency parametrization (SURY) [*Wouters et al.*, 2016] and the coupling with the turbulence kinetic energy based surface-layer transfer module of the COSMO(-CLM) model [*Doms et al.*, 2011].

Text S2: Downscaling setup. The configuration of the climate downscaling for 1980-2014 used in this study is based on the Convection-Resolving Climate Setup (CRCS) developed by the CLM-community and adapted for the Belgian extent [*Brisson et al.*, 2016a, b]. The domain has a size of 538 km by 490 km (192 by 175 grid cells) centered over Brussels with a horizontal grid spacing of 2.8 km and 40 vertical layers with the lowest level at 10 m above the ground. The analysis subdomain covers 95 by 95 grid cells (266 km by 266 km) encompassing Belgium, hence providing a boundary buffer zone of more than 100 km at each side. The downscaling takes lateral boundary conditions from the ERA-INTERIM-driven COSMO-CLM simulation at 12.5 km resolution in the COordinated Regional climate Downscaling EXperiment (CORDEX) for Europe [*Kotlarski et al.*, 2014; *Jacob et al.*, 2013; *Vautard et al.*, 2013].

Land-surface parameters for soil type, vegetation and orography are specified with the External Parameter tool (EXTPAR) [*Smiatek et al.*, 2008]. These are processed from global land-cover data sets for orography and coastlines from ASTER (Three Arc-Second Elevation dataset), soil data from the 'Digital Soil Map of the World' (DSMW; see `http://data.fao.org/map?entryId=446ed430-8383-11db-b9b2-000d939bc5d8`), and land-use data from CORINE land-cover for the year 2000. The vegetation parameters, which include vegetation cover fraction, Leaf-Area-Index and rooting depth, are specified with annual minimum and maximum values depending on the land-use [*Doms et al.*, 2011, see

their tables 14.3, 14.4, and 14.5]. For the extra-tropical Northern Hemisphere, a growing and resting period is calculated according to the latitude. The roughness length parameter (over land) depends on both land-use and the subgrid-scale orography.

The latest release of the EXTPAR tool has been extended with the additional surface fields required by TERRA_URB, which include the Impervious Surface Area (ISA) and annual-mean Anthropogenic Heat Flux (AHF). The ISA-field refers to the abundance of buildings, streets, parkings and other man-made water-impermeable pavements), and it is derived from the land-use modelling scenarios described in the Text S3. The AHF-field is obtained from a global dataset [*Flanner*, 2009], which is generated from country-specific data of energy consumption from non-renewable sources. This was apportioned according to population density (conserving the national total) and converted to annual-mean gridded energy flux at a resolution of 2.5×2.5 minutes. The annual-mean values are further redistributed to the native model grid at 2.8 km resolution according to the ISA-field. Hereby, it is assumed that areas with large ISA fraction (including industrial areas with low population densities) have higher anthropogenic heat emissions. For Brussels, the AHF reaches an annual-mean value of $49.16\,\mathrm{W\,m^{-2}}$ in the city centre. These values are of the same magnitude to that obtained by two other methodologies presented in [*Van Weverberg et al.*, 2008] yielding $43.8\,\mathrm{W\,m^{-2}}$ and $46.7\,\mathrm{W\,m^{-2}}$. While not considering this as a formal validation, the similarity of magnitude of the results obtained by three very different methods inspires confidence.

Text S3: Determination of the impervious surface area. The ISA-field is derived for the historical (LND:HST) and business-as-usual (LND:BAU) land-use scenarios described in the Materials and Methods. This is done under the assumption of a

relationship between ISA on the one hand and land use and the urbanization rate on the other hand. The urbanization rate is calculated from the CORINE land-use map for 2000 according to the urban and industrial land use in a 1.5 km radius around each grid-point at its native resolution (300 m resolution) and results in a map showing the percentage of urbanized land in the neighborhood of each cell. The resulting urbanization map for the year 2000 is divided into ten equal interval categories, ie., 0-10%, 10-20%, 20-30%, etc. In a second step, an overlay of this categorized map with the land-use map is made. This results in a map of 233 possible combinations (24 land-use types with the 10 urban categories where 7 combinations did not occur). In a next step, an overlay is made with the ISA extracted from the soil-sealing map brought by the European Environmental Agency [*Maucha et al.*, 2010], see `http://www.eea.europa.eu/articles/urban-soil-sealing-in-europe`. Hereby, the averaged value of the ISA for each of the combinations is recorded. Finally, these values are applied to the original land-use map of 2000 (LND:HST). Upscaling to the native grid of the regional climate model over Belgium with a horizontal resolution of 2.8 km, the method has a very good skill in reproducing the original ISA dataset (see Figure S6). A Pearson correlation coefficient of 0.93 is found with a slight overestimation in the rural areas and a slight underestimation in the dense urban areas. Finally, the average value calculated for the year 2000 for each combination of land use and category of the urbanization rate is assumed to be valid in 2060, and is applied to the simulated land-use and urbanization map for 2060 (LND:BAU).

Text S4: Model evaluation. The COSMO-CLM model configured at the convection-permitting scale for Belgium shows a very good skill in the representation of precipitation and cloud properties from in-situ and satellite observations [*Brisson et al.*, 2016b, a].

Moreover by the coupling to its urban land-surface scheme TERRA_URB, it was able to reproduce very well the temperature variability in mid-latitude urban areas according to in-situ observations, satellite imagery and tower profiles [*Wouters et al.*, 2016; *Trusilova et al.*, 2016]. TERRA_URB also matches very well the observed components of the urban surface energy balance from intensive urban observation campaigns for Basel, Toulouse and Singapore [*Wouters et al.*, 2015; *Demuzere et al.*, 2017].

An evaluation of two-metre air temperature (taken as input for the urban heat stress indicator) is performed against observations in the study domain. First, the model's capability of the short-term trends in city-scale temperature variability on an hourly basis - particularly the magnitude, day-to-day variability and diurnal cycle of the urban heat island intensity - is evaluated by comparing with the urban climate observations from *De Ridder et al.* [2015]. Afterwards, the model skill in long-term climatological trends of the coarse-scale temperature variability on a daily basis is evaluated against temperature fields from the E-OBS gridded climatological dataset [*Haylock et al.*, 2008] with a spatial resolution of 25 km. In addition to the evaluation for air temperature given below, previous studies have shown that the COSMO-CLM model shows a good skill in other meteorological parameters than two-metre temperature as well for our region of interest, including precipitation, cloud properties and urban land-surface temperatures from in-situ and satellite observations, and urban boundary-layer temperature profiles from tower observations [*Brisson et al.*, 2016a, b; *Wouters et al.*, 2016]. Furthermore,

the urban climate modelling methodology with TERRA_URB has been evaluated against intensive urban observation campaigns for cities (Berlin, Basel, Toulouse and Singapore) outside of the study domain as well [*Demuzere et al.*, 2017; *Trusilova et al.*, 2016; *Wouters et al.*, 2015]. This includes the evaluation of the different components of the urban surface-energy balance that drive the urban heat island effect.

Comparison with urban-climate observations. The urban climate-observations consist of records from urban/rural site pairs for the cities Antwerp and Ghent, which started in 2012. The coordinates of the urban/rural site pairs for Antwerp and Ghent correspond to those listed in Tab. S1, and can also be found on the map in Figure S1. A model evaluation against these measurements is performed for the overlapping records between model and observations (years 2012-2014) during the extended summer months (April-September), see Figure S2. The urban heat islands (calculated as the hourly temperature difference from the urban/rural site pairs) are well reproduced with a mean bias of -0.24 (-0.06) and a Pearson Correlation Coefficient of 0.65 (0.68) for Antwerp (Ghent). Focusing further on Antwerp during a mid-summer period in 2012 containing a heatwave (see Figure S3), the climate model agrees very well with the observed temperature variability for the urban site with a Mean Bias (MB) of 0.39 K and a Pearson Correlation Coefficient (PCC) of 0.96. This is also the case for the rural site with a MB of 0.56 K and a PCC of 0.95. The positive bias generally stems from the overestimation of nocturnal temperatures, which is consistent with the results of the minimum temperatures in the coarse-scale climatology discussed in more detail in the next paragraph. The model can reproduce the evolution of the urban heat island with a PCC of 0.80. Especially, the representation of the day-to-day variability and the diurnal cycle are very well represented

in the model. It also captures very well the timing of the maximum in daily-averaged UHI intensities, which occurs during the warm period (25^{th} - 28^{th}) July and the heat-wave period (17^{th} - 19^{th} August). The modelled averaged UHI intensity (1.56 K) matches very well the observed intensity (1.74 K) with a MB of -0.18 K. The underestimation results from the negative bias in the nocturnal UHI. In turn, it stems from a positive bias in the nocturnal temperatures in the rural area. The current model setup has a very good performance in reproducing the day- and night-time UHI magnitudes with respect to the existing capability of state-of-the-art convection-permitting models (CPM), as benchmarked for mid-latitude European cities in Table S4.

Comparison with coarse-scale climatological observations. The model is capable of capturing the coarse-scale temperature evolution from E-OBS. Particularly, a root-mean-square error (RMSE) of the monthly timeseries in the minimum (maximum) temperatures over the Belgian domain amounts to 1.52 °C (1.10 °C). Such a skill score is excellent compared to a state-of-the-art multi-model evaluation [*Davin et al.*, 2016] against E-OBS, yielding an RMSE-range between 1.3 and 2.3 °C (between 1.3 and 2.5 °C). The model also can also reproduce the coarse-scale spatial variation in the climatological averages of the temperatures in the observations, see Figure S4. Particularly, the coastal-inland gradient in the observed maximum temperatures, with lower values in the coastal areas compared to the inland areas, is well represented by the model. Both the model and observations show a maximum in the domain center around Brussels for the minimum temperature, and additional hot pools in the Northeast and the South for the maximum temperature. They also show a cold pool over the hills in the Southeast of the domain for both minimum and maximum temperature. The annual model bias for the daily min-

imum and maximum temperatures amounts to 1.04 °C and −0.52 °C, respectively. The overestimation of the minimum temperatures is commonly found in climate models [*Brisson et al.*, 2016a; *Davy and Esau*, 2014; *Jacob et al.*, 2007] which may be related to the overestimation of turbulent mixing rates for stably-stratified (nocturnal) regimes in existing boundary-layer schemes [*Dimitrova et al.*, 2016]. A part of the positive bias in the minimum temperatures may also be related to the fact that urban heat islands, reaching their maximum during night-time, are resolved by the convection-permitting simulations, whereas such local-scale features are only partially represented - or even missing - in the E-OBS dataset. The latter is due to the fact that most of the underlying stations follow the WMO observation standards [*WMO*, 2008], for which the observations take place over a grassland away from obstacles such as buildings or trees. In order to be representative for the surrounding area, the stations are also located outside of the city centers.

Text S5: Bias correction. Even though the model has very good skill compared to other models for reproducing the small-scale urban-heat islands and the coarse-scale temperature climatology (see previous section), the heat-stress indicator is very sensitive to the model bias. Therefore, a coarse-scale bias correction is applied according to the gridded daily temperature fields of E-OBS [*Haylock et al.*, 2008] that preserves the local-scale features such as the urban heat islands resolved by the CPS downscaling. For each domain gridcell, bias-corrected daily values (minima and maxima) of two-meter temperature at domain position (i, j) are calculated:

$$T^{\mathrm{BC}}_{i,j} = T^{\mathrm{M}}_{i,j} + \left[\left(T^{\mathrm{O}} - T^{\mathrm{M}}\right) * G\right]_{i,j} + \mathrm{UHIBC}_{i,j} \tag{1}$$

where the second right-hand term represents the model bias at the coarse spatial scales, and the last right-hand term the UHI bias correction (specified further). Hereby, '$*$' refers

to the convolution operator in space of the model bias and a two-dimensional Gaussian kernel function G with a standard deviation of 25 km that is equal to the resolution of the E-OBS dataset. Such a formulation is valid for the natural environments where the WMO measurements underlying the E-OBS dataset take place (see also Text S4). However, it fails for urban areas as the underlying measurements generally avoid local climate characteristics such as urban heat islands (see Text S4). Therefore, the formulation is modified by disregarding urban heat islands in determining the coarse-scale bias:

$$T^{\mathrm{BC}}_{i,j} = T^{\mathrm{M}}_{i,j} + \frac{\left[\left(\mathbb{1}_{\mathrm{R}} \odot \left[T^{\mathrm{O}} - T^{\mathrm{M}}\right]\right) * G\right]_{i,j}}{\left[\mathbb{1}_{\mathrm{R}} * G\right]_{i,j}} + \mathrm{UHIBC}_{i,j} \tag{2}$$

where '$\odot$' represents the point-wise matrix product, and '$\mathbb{1}_{\mathrm{R}}$' the one-matrix that is unity at rural locations and zero at urban areas. In the latter, a gridcell is considered rural if the ISA of the surrounding 100 km^2 is less than 25%. The UHI bias correction is calculated as follows:

$$\mathrm{UHIBC}_{i,j} = c\ \mathrm{ISA}\left(T^{\mathrm{M}}_{i,j} - \left[\left(\mathbb{1}_{\mathrm{R}} \odot T^{\mathrm{M}}\right) * G\right]_{i,j}\right) \tag{3}$$

The term in the outer brackets represent the urban heat island intensity, $c = (1 - 1/(1 - 0.25)) = 0.33$ is the factor compensating for an underestimation of the observed canopy-layer UHI in the minimum temperatures of 25%, and ISA the Impervious Surface Area. By employing the bias correction, the modelled record of the heat-stress indicator matches very well the urban-climate observations in Antwerp during the years 2012, 2013 and 2014, especially the urban/rural contrast (see Figure S5).

For the CPS climate downscaling assuming the land-use scenario for the reference year 2060 (see Section 2.3 of the main text), the bias correction for each time step is adopted

from that of the reference year 2000 calculated above:

$$T_{i,j}^{\mathrm{BC},2060} = T_{i,j}^{\mathrm{M},2060} + T_{i,j}^{\mathrm{BC}} - T_{i,j}^{\mathrm{M}} \tag{4}$$

Text S6: Motivation of the heat stress indicator. The heatwave thresholds used in the heat-stress indicator (see section 2.2) are in agreement with the local governmental heatwaves-alarm levels provided by the Belgian Federal Public Service for Health, Food Chain Safety and Environment [*Belgian Government*, 2017]. Such an approach is recommended, because there is no common globally-accepted heatwave definition available yet [*Xu et al.*, 2016]. Particularly, the relation between temperatures and excessive mortality rates depends on the region under scope [*Baccini et al.*, 2008], hence supporting the usage of a country-specific heat-stress indicator. The temperature thresholds for minimum and maximum temperature were obtained from the 95^{th} percentile of the summer temperatures (June - August) of the Belgian climatology between 1970 and 2004. The application of percentile-based thresholds is recently supported and employed by existing studies about the attribution of excessive temperatures to heatwave-related problems [*Tong et al.*, 2015; *Rey et al.*, 2007], and it's also consistent with previous heatwave assessments [eg., *Fischer et al.*, 2012; *Meehl and Tebaldi*, 2004; *Huth et al.*, 2000]. The usage of both minimum and maximum temperatures in the heatwave definition is justified by the fact that there is no clear superior predictive capacity among the different temperature quantities to heatwave-related problems such as excessive mortality rates [*Hondula et al.*, 2014; *Barnett et al.*, 2010].

The heatwave definition takes temperature into account whereas other meteorological variables (such as moisture, wind speed and radiation) are omitted. This is in line with a large body of literature investigating the occurrence of heatwave-related problems (as

listed in the introduction), which consider temperature as a predictor, yet do not consider the other meteorological variables. Further support is provided by *Barnett et al.* [2010] who did not find a superior predictive capacity of more complex (humidity-based) metrics compared to temperature-only metrics for heat-related mortalities.

Text S7: Motivation and interpretation of the global emission scenarios. By combining the different RCP scenarios into the global emission scenarios (instead of considering RCP scenarios separately), the sample size becomes large enough to provide robust quantiles of the climate-change signals. The average change in daily minimum and maximum temperature over the extended summer months (April-September; used for the heat-stress indicator) according to the best-case scenario (EMI:BEC) is comparable to that according to the standard deviation of RCP2.6 (see Figure S7, panel C), whereas the change according to the worst-case scenario (EMI:WOC) is comparable to that of the upper standard deviation of RCP8.5. The average change according to the median scenario (EMI:MED) is comparable to that between RCP4.5 and RCP6.0.

Because of the limitations in today's computational infrastructure, urban CPM ensemble datasets are not available yet. In order to provide a dataset large enough, the above-mentioned CMIP5 ensemble dataset is used. These above global emission scenarios only consider emission-induced changes in the climate feedbacks (eg., between temperature, moisture and wind) at the global- and synoptic-scale, whereas those changes at the local/city-scale are neglected.

Text S8: Assessment limitations and perspectives

S8.1: Heat-stress assessment. Heat-stress assessments and projections could lead to different outcomes when other heat-stress metrics or models are used [*Xu et al.*, 2016;

Buzan et al., 2015; *Hondula et al.*, 2014]. For example, urban heat islands tend to reach their maximum (minimum) during night-time (daytime) [*Wouters et al.*, 2016, 2013; *Bohnenstengel et al.*, 2011], hence the relative role of urbanization in the current assessment would become larger (smaller) when only considering nocturnal (daytime) temperatures, as found earlier by *Hondula et al.* [2014] for Maricopa County (Arizona). These different outcomes are despite the fact that there is no superior predictive capacity to heatwave-related problems among different exposure metrics such as excessive mortality rates [*Hondula et al.*, 2014; *Barnett et al.*, 2010]. There have already been many existing studies about the occurrence of heatwave-related problems (see introduction), and many heat-stress indices have been developed for different applications as well [*Buzan et al.*, 2015; *Matzarakis et al.*, 2010; *Epstein and Moran*, 2006]. Still, optimal choices for their meteorological predictors (taking into account minimum, mean and maximum temperature, humidity, wind speed, radiation...) are largely unaddressed, and they require future research towards a universal heatwave definition. In this respect, it needs to be noted that there has been released a first conclusive study on this topic by *Mora et al.* [2017]: based on a literature review of 683 cases of excessive heat-related mortality, they found that temperature and relative humidity most accurately distinguished between past lethal and non-lethal heat episodes compared to wind speed, radiation and several other metrics. Yet, future research should also address the more appropriate metrics for other heatwave-related problems than heat-related mortality as listed in the introduction (eg., for energy consumption). Considering that a higher humidity reduces the ability of the human body to cool down by transpiration, it is expected that the excess heat stress by urbanization during daytime will be tempered because of a deficit in evapotranspiration (hence lower

humidity), whereas during nighttime it will remain similar because of negligible evaporation (hence similar humidity) in both urban and rural areas. As the current study involves an assessment with a heat-stress indicator considering local heatwave alarm levels based on temperature only, the optimal choice and sensitivity to the heat-stress metric is not investigated further.

S8.2: Urban-climate observations, modelling and projections. The delta-change approach in the current study does not consider greenhouse-gas induced changes of the local (meso- and city-scale) land-atmosphere feedbacks in urban areas. Hence, their additional influence on urban heat stress under climate change still needs to be addressed. They may e.g. involve changes in the local interplay between (soil) moisture, heat storage, circulations (sea breeze, city breeze...) and radiation during the climate extremes for the future, which may further modify future urban heat stress. These changes could occur in addition to the excess heat stress under climate change established by large-scale greenhouse-gas induced climate change and local land-use change (urban expansion) addressed in our study. Assessing the greenhouse-gas induced local feedbacks requires the downscaling of multiple GCM ensemble members with urban CPM. Noting the considerable computational cost of CPM, this challenge - together with the need to downscale other regions - require new strategies of dynamical and statistical model techniques, including the downscaling with GPU-based hardware accelerators [*Leutwyler et al.*, 2016]. Finally, extisting datasets of urban-climate observations are only limited in time. They should be maintained and extended in order to evaluate the climatological UHI trends of urban-climate models.

S8.3: Land-use modelling. The effects established by transient local land-use are not considered in the current assessment, hence the community should aim for CPM simulations with an evolving landscape. The community should also aim for local land-use and adaptation scenarios that are in concert with the global emission scenarios. Such an approach would allow to evaluate the reinforcing impacts of urbanization pathways (eg., the consideration of compact city design) at the different scales.

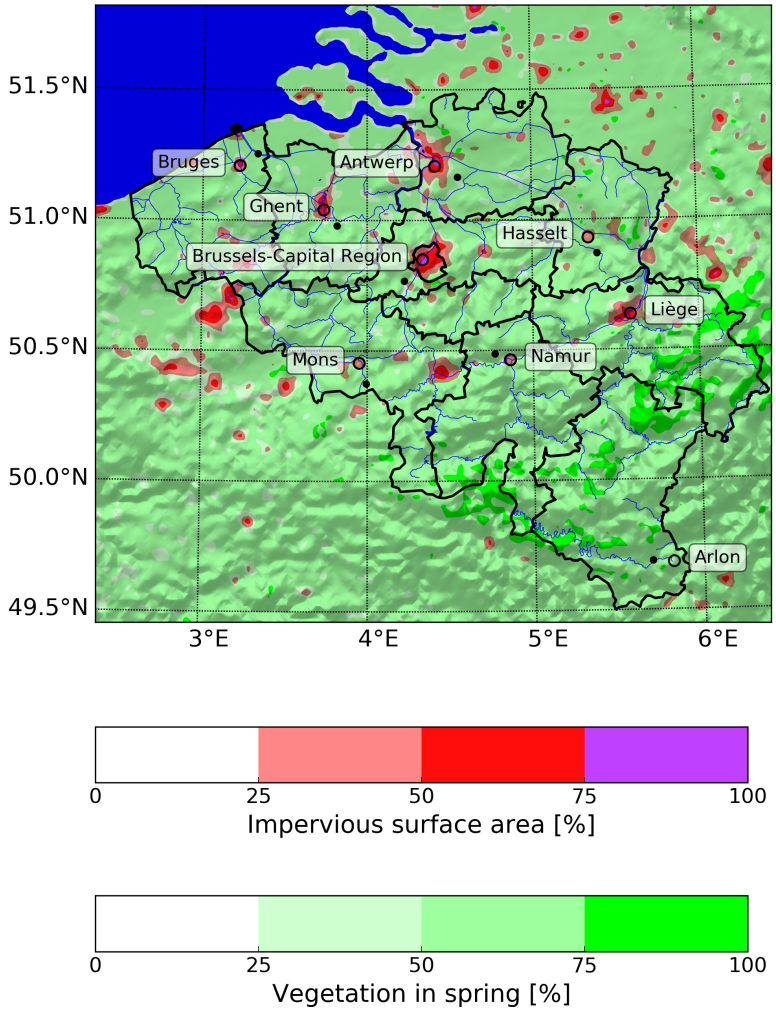

Figure S1. Land-surface parameters and topography (shadow mask) from the convection-permitting downscaling with the COSMO-CLM model for the analysis subdomain (266 km by 266 km) encompassing Belgium. The locations of the urban site (circles) and the rural site (bullets) for the Brussels-Capital Region and the provincial capital cities in Belgium (listed in Tab. S1) are indicated on the map.

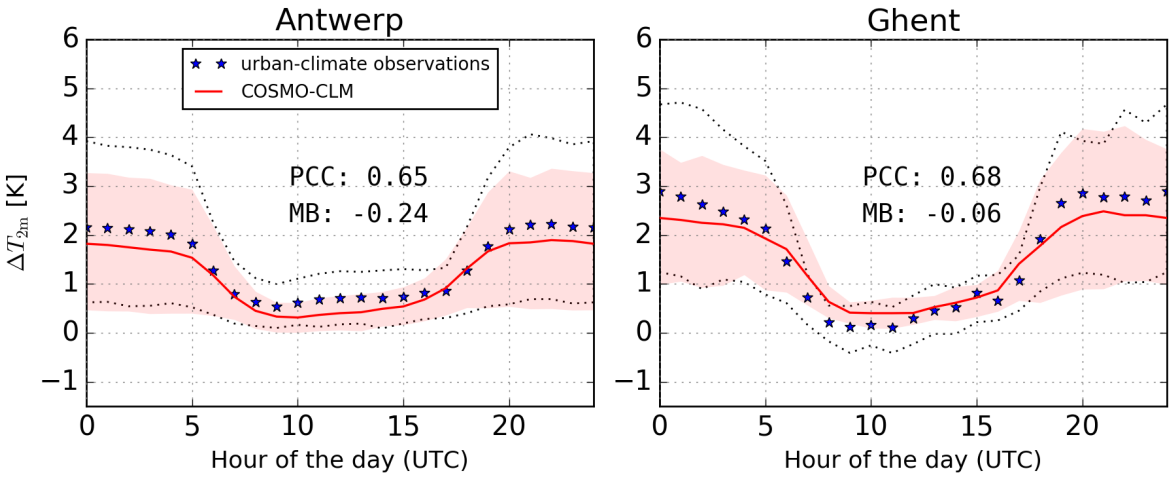

Figure S2. Mean diurnal cycle of the urban heat island for the urban-climate observations (stars) and the COSMO-CLM model (full lines) available during the extended summer periods (April-September) for the years 2012 to 2014 for Antwerp (left) and Ghent (right). The urban heat island for Antwerp (Ghent) is calculated from the temperature difference between the urban station at Royal Lyceum in Antwerp (at Vrijdagmarkt in Ghent) and the rural station at Organic Farm Van Leemputten in Vremde (at Proefhoeve in Melle). The dotted lines (red shaded areas) indicate the range between the 16^{th} and 84^{th} percentile values (percentile levels agree with one standard deviation for a normal distribution) of the observed (modelled) temperatures. The coordinates of the urban and rural location correspond to those listed in Table S1 for Antwerp and Ghent, and can also be found on the map in Figure S1. The Pearson Correlation Coefficient (PCC) and the mean bias (MB) of the underlying hourly timeseries are annotated in each panel.

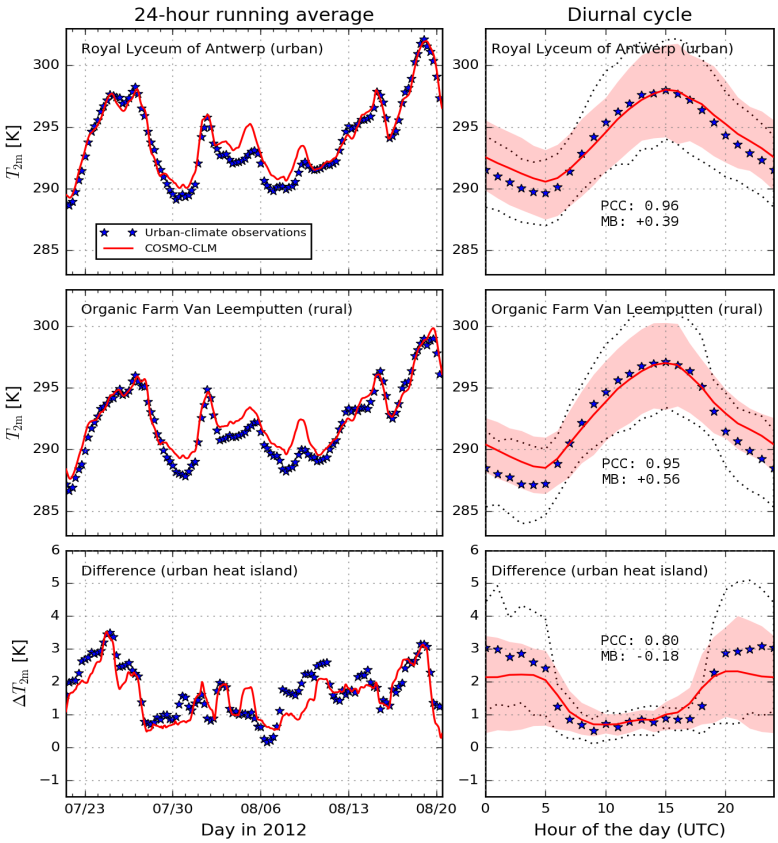

Figure S3. 24-hour running averages and mean diurnal cycle for temperatures according to the urban-climate observations (stars) and the COSMO-CLM model (full lines) during mid-summer (2012/07/21 until 2012/08/20) at the Royal Lyceum in Antwerp (urban), the Organic Farm Van Leemputten in Vermde (rural), and their difference (urban heat island effect of Antwerp). The dotted lines (red shaded areas) indicate the range between the 16$^{\text{th}}$ and 84$^{\text{th}}$ percentile values (percentile levels agree with one standard deviation for a normal distribution) of the observed (modelled) temperatures. The coordinates of the urban and rural location correspond to those listed in Table S1 for Antwerp, and can also be found on the map in Figure S1. The Pearson Correlation Coefficient (PCC) and the mean bias (MB) of the underlying hourly timeseries are annotated in each right panel.

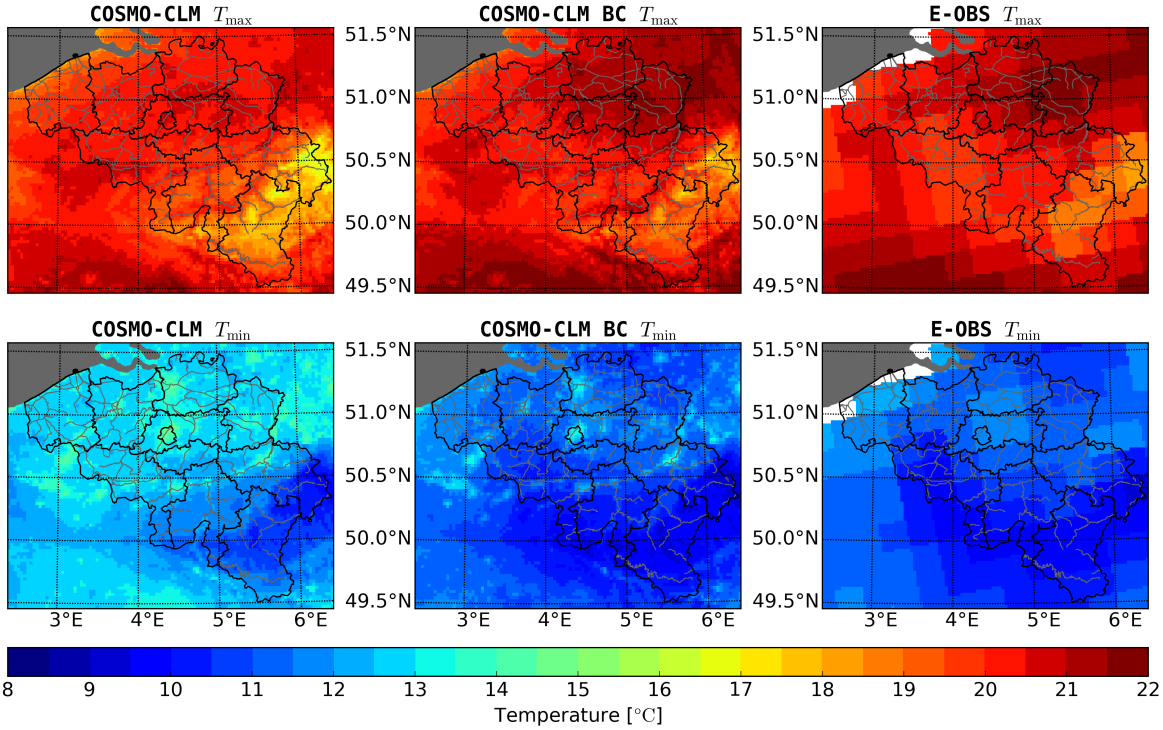

Figure S4. Climatological averages (1980-2014) for daily maximum (top row) and minimum (bottom row) temperatures. The results are shown for the extended summer months (beginning of April until the end of September) consistent with the yearly timespan required for the heat stress indicator (see materials and methods). The left, middle and right panels represent the model output (COSMO-CLM), the bias-corrected model output (COSMO-CLM BC), and the gridded dataset (E-OBS), respectively.

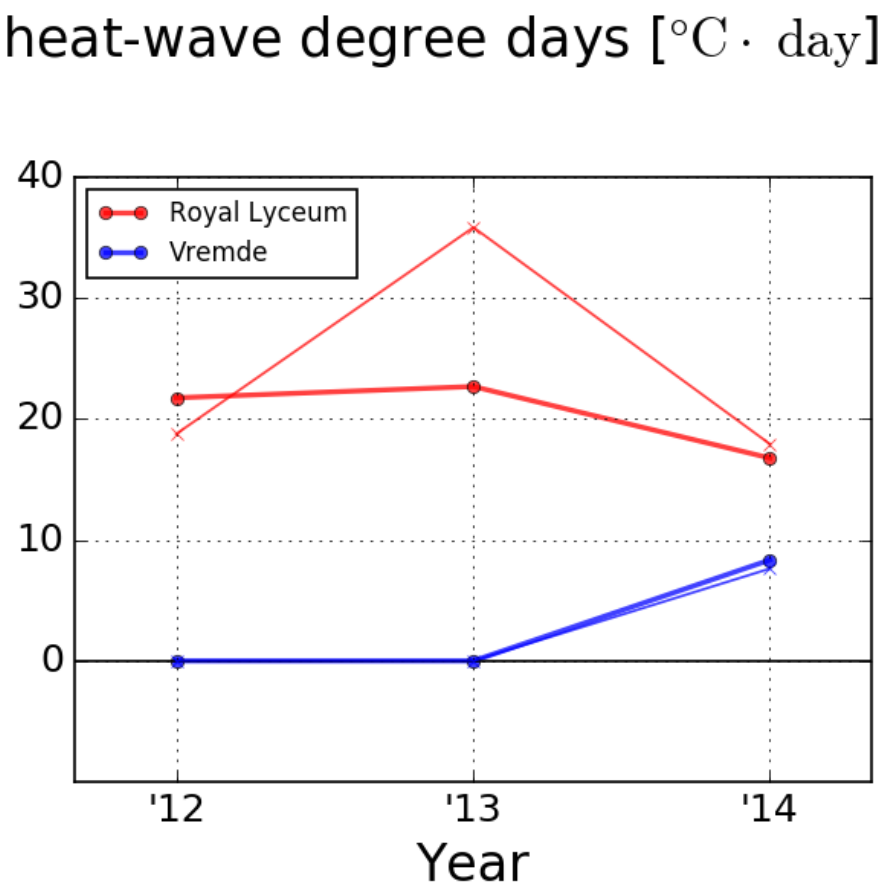

Figure S5. The heat stress indicator according to observed and modelled temperature records during the urban-climate observation campaign for Antwerp between 2012 and 2014. The observations (thin lines) are obtained from in-situ urban climate observations, whereas the modelled values (thick lines) correspond to the bias-corrected climate downscaling. The heat stress for the urban (red) and the rural (blue) site is shown, indicating the differential heat stress established by the urban climate of Antwerp. The coordinates for the site pair correspond to those listed in Table S1 for Antwerp, and can also be found on the map in Figure S1.

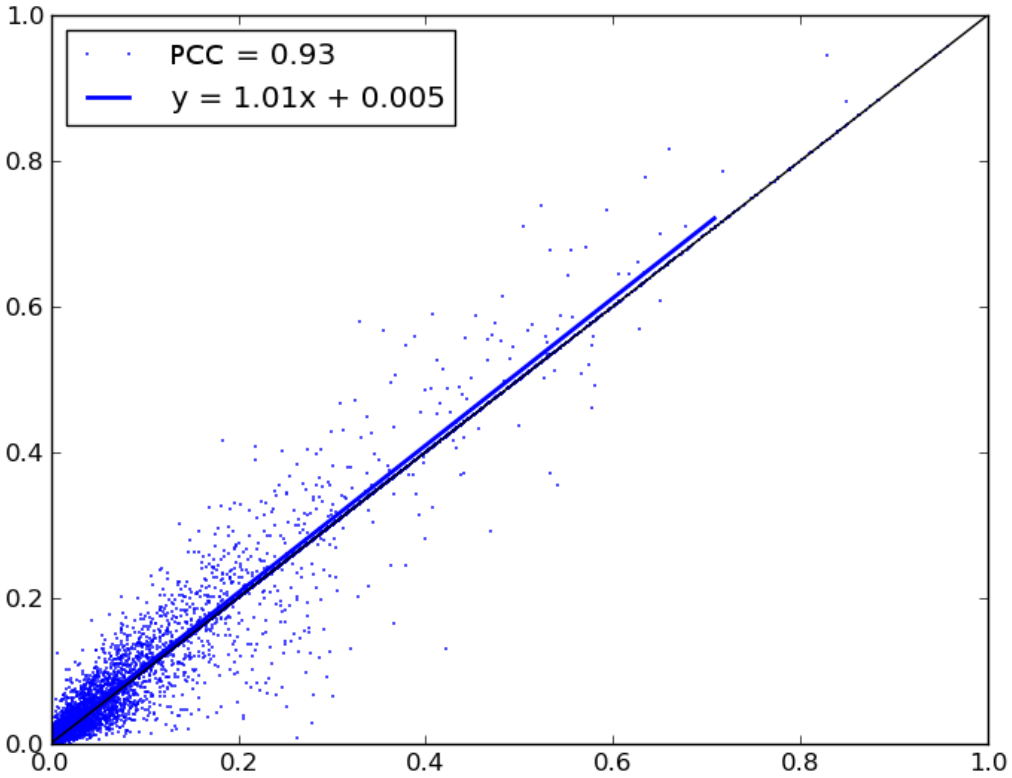

Figure S6. Correspondence between original impervious surface area and that derived from the spatial model (see Materials and Methods) for the reference year 2000. Each of the blue dots refers to one grid cell within Belgium of the horizontal atmospheric model domain at 2.8 km resolution. The values on the X-axis represent the impervious surface area upscaled from the soil-sealing map brought by the European Environmental Agency [*Maucha et al.*, 2010], and the values on the Y-axis those derived from the land-use model for the reference year 2000.

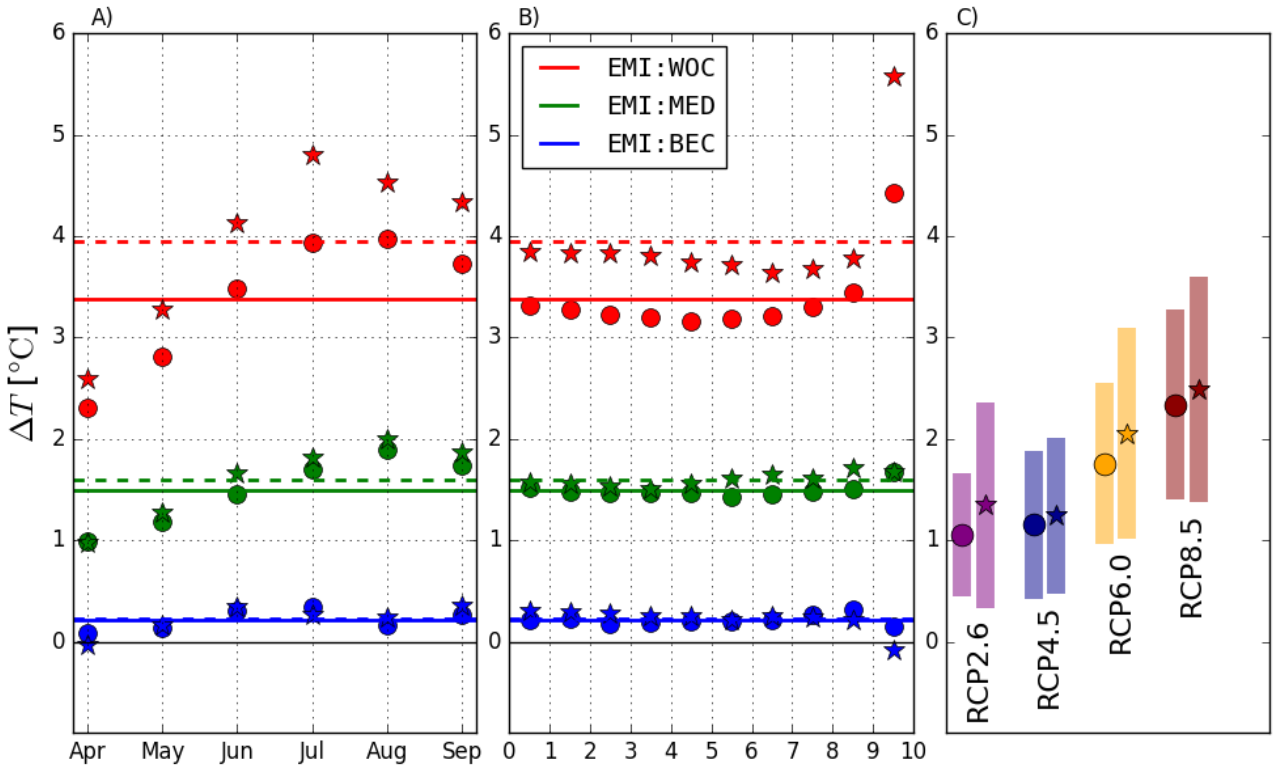

Figure S7. Panels A and B show the changes in the daily minimum and maximum temperature between the periods 1980-2014 and 2040-2074 according the three global emission scenarios EMI:BEC (best-case), EMI:MED (median) and EMI:WOC (worst-case) for Belgium, which are described in Section 2.4 of the main text. The values averaged for each summer month are given with the markers on panel A (bullets for minimum temperature; stars for maximum temperature), whereas the values for each 10-category temperature quantile bin averaged over the extended summer months (April-September) are given on panel B. The full lines (dashed lines) in both panels show the summer-averaged change in the minimum (maximum) temperatures. The summer-average changes (bullets for minimum temperature; stars for maximum temperature) and standard deviation (bars) for the distinct RCP ensemble climate scenarios are provided on panel C.

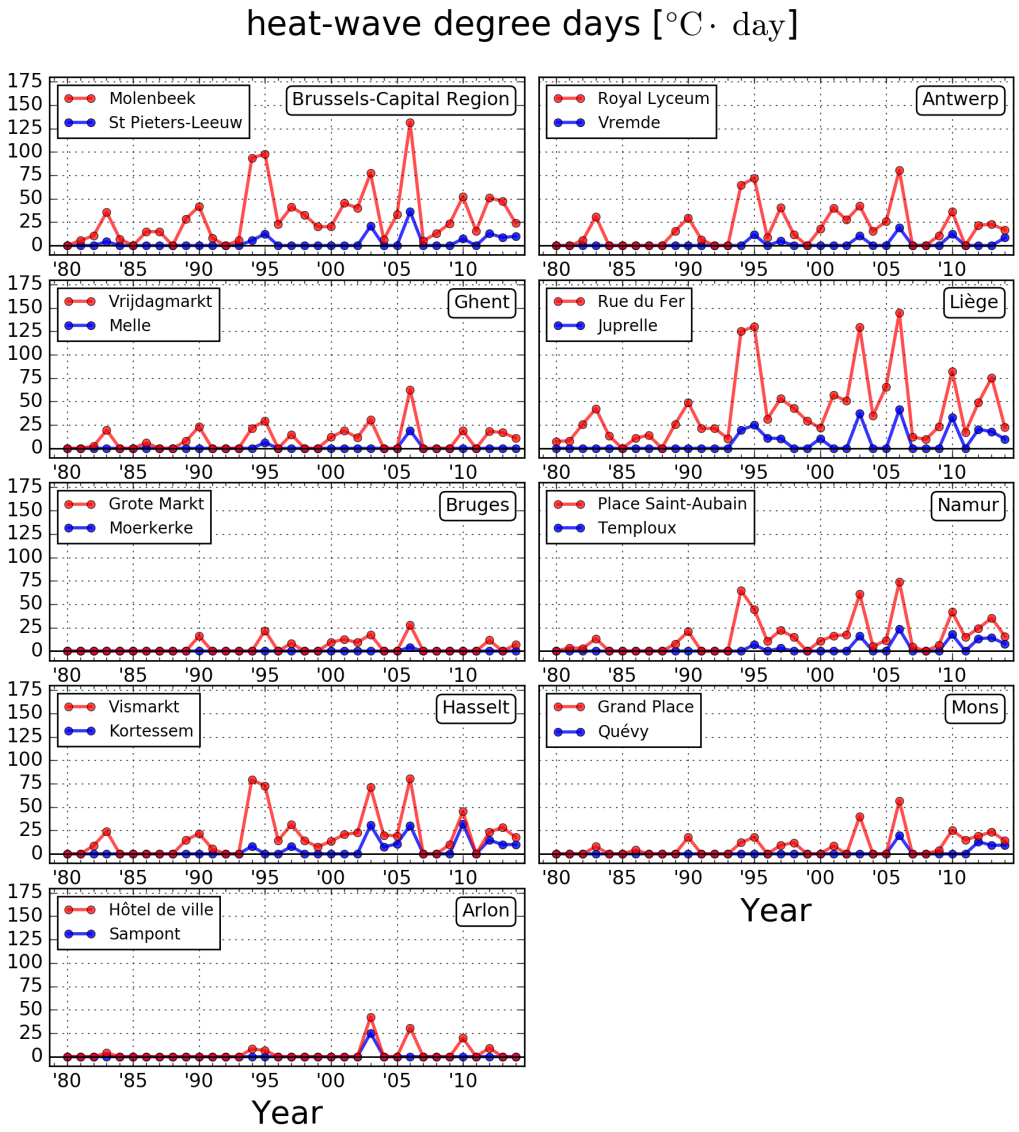

Figure S8. Historical climate downscaling of urban heat stress for the Brussels-Capital Region and the provincial capital cities in Belgium. For each location, the heat stress for the city center (red) and for a neighboring rural site (blue) is shown, indicating the differential heat stress established by the urban climate. The heat stress is expressed in heatwave degree days according to the heat-stress indicator provided in Section 2.2 using output from the climate downscaling presented in Section 2.1. Locations are indicated on the map in Figure S1 and the (cities') characteristics are listed in Table S1.

City	Size	Imperviousness	Elev.	Urban site	Lat. [°]	Lon. [°]	Rural site	Lat. [°]	Lon. [°]
Brussels-Capital Region	1200	0.61	31	Molenbeek	50.8500	4.3340	St Pieters-Leeuw	50.7680	4.2240
Antwerp	517	0.52	6	Royal Lyceum	51.2085	4.4102	Vremde	51.1660	4.5490
Ghent	257	0.43	4	Vrijdagmarkt	51.0410	3.7350	Melle	50.9800	3.8160
Liège	200	0.42	77	Rue du Fer	50.6351	5.5860	Juprelle	50.7276	5.5876
Bruges	117	0.25	2	Grote Markt	51.2090	3.2240	Moerkerke	51.2550	3.3630
Namur	110	0.22	105	Place Saint-Aubain	50.4642	4.8618	Temploux	50.4870	4.7691
Hasselt	75	0.22	43	Vismarkt	50.9300	5.3390	Kortessem	50.8695	5.3917
Mons	95	0.19	39	Grand Place	50.4540	3.9510	QuÃľvy	50.3743	3.9949
Arlon	29	0.06	386	Hôtel de ville	49.6833	5.8167	Sampont	49.6875	5.6934

Table S1. Characteristics of the Brussels-Capital Region and the provincial capital cities in Belgium. The city size is expressed in thousands of people and the elevation in meters above sea level. The city imperviousness is expressed as the impervious surface area (ISA) fraction for the 100 km^2 around the city center. For each city, coordinates of the urban/rural site pairs are given for which the heat-stress indicator is calculated.

ID	Emission scenario	(period)	Land-use scenario	(ref. year)
EMI:HST LND:HST	historical	(1980-2014)	historical	(2000)
EMI:HST LND:BAU	historical	(1980-2014)	future	(2060)
EMI:BEC LND:HST	5th percentile	(2040-2074)	historical	(2000)
EMI:BEC LND:BAU	5th percentile	(2040-2074)	future	(2060)
EMI:MED LND:HST	50th percentile	(2040-2074)	historical	(2000)
EMI:MED LND:BAU	50th percentile	(2040-2074)	future	(2060)
EMI:WOC LND:HST	95th percentile	(2040-2074)	historical	(2000)
EMI:WOC LND:BAU	95th percentile	(2040-2074)	future	(2060)

Table S2. Overview of historical climate downscaling and projections based on ensemble-based EMIssion scenarios (EMI) and LaND-use scenarios (LND). The emission scenarios are constructed by applying ensemble climate-change statistics from general circulation models on the monthly downscaled temperatures of CMIP5 [*Taylor et al.*, 2012]. Hereby, the change in daily minimum and maximum temperatures are considered with respect to predefined ensemble percentiles resulting in three emission scenarios, namely the best-case (BEC; 5th percentile), median (MED; 50th percentile) and worst-case (WOC; 95th percentile) scenario. A distinction is made among the changes in the 10-category quantiles of the temperature distribution in each month. Two land-use scenarios are considered, namely a historical scenario for the year 2000 (LND:HST) and a future scenario for the year 2060 (LND:BAU) according to a business-as-usual land-use change [*Acosta-Michlik et al.*, 2011]. These land-use change scenarios are based on the output of a spatial model, which is taken into account in the land-surface parameters by the climate downscaling. Details about the emission scenarios and land-use change scenarios can be found in the methods section of the main text.

Model	Research centre	Resolution (Lon × Lat)	Scenario			
			RCP2.6	RCP4.5	RCP6.0	RCP8.5
BNU-ESM	College of Global Change and Earth System Science, Beijing Normal University, China	$\sim 2.8° \times 2.8°$	r1i1p1	r1i1p1		
CNRM-CM5	Centre National de Recherches Météorologiques/Centre Européen de Recherche et Formation Avancées en Calcul Scientifique, France	$\sim 2.8° \times 2.8°$		r1i1p1		r1i1p1
GFDL-CM3	NOAA Geophysical Fluid Dynamics Laboratory, USA	$2.5° \times 2.0°$	r1i1p1		r1i1p1	r1i1p1
GFDL-ESM2G	NOAA Geophysical Fluid Dynamics Laboratory, USA	$2.5° \times 2.0°$	r1i1p1	r1i1p1	r1i1p1	r1i1p1
GFDL-ESM2M	NOAA Geophysical Fluid Dynamics Laboratory, USA	$2.5° \times 2°$		r1i1p1	r1i1p1	r1i1p1
inmcm4	Institute for Numerical Mathematics, Russia	$2° \times 1.5°$		r1i1p1		
IPSL-CM5A-LR	Institut Pierre-Simon Laplace, France	$3.75° \times 1.875°$	r3i1p1			r3i1p1
PSL-CM5A-MR	Institut Pierre-Simon Laplace, France	$2.5° \times 1.25°$	r1i1p1		r1i1p1	r1i1p1
MIROC-ESM	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies, Japan	$\sim 2.8° \times 2.8°$	r1i1p1	r1i1p1	r1i1p1	r1i1p1
MIROC-ESM-CHEM	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies, Japan	$\sim 2.8° \times 2.8°$	r1i1p1	r1i1p1	r1i1p1	r1i1p1
MRI-CGCM3	Meteorological Research Institute, Japan	$1.125° \times 1.125°$	r1i1p1	r1i1p1		r1i1p1

Table S3. Summary of the 31 ensemble members from the Coupled Model Intercomparison Project of the World Climate Research Programme - Phase 5 (CMIP5) multi-model experiment [*Taylor et al.*, 2012], considering the different Representative Concentration Pathways [*Moss et al.*, 2008]. Besides greenhouse gas emissions, the ensemble also takes into account other anthropogenic forcings at the global scale including aerosols and global land-cover change.

City [location] (study) Convection-permitting model Period	urban land-surface schemes	UHI day			UHI night		
		observed	modelled	bias	observed	modelled	bias
Antwerp [Royal Lyceum] (current) COSMO-CLM 2012/07/21 - 2012/08/21	TERRA_URBv2.0	0.8	1.0	0.2 (25%)	2.7	2.2	-0.5 (-19%)
Berlin [Alexanderplatz] [*Trusilova et al.*, 2016] COSMO-CLM 2012/07/01 - 2012/08/01	TERRA_URBv1.0 TEB DCEP	1.0	1.8 1.7 1.6	0.8 (80%) 0.7 (70%) 0.6 (60%)	2.0	2.0 1.8 2.3	0.0 (0%) -0.2 (-10%) 0.3 (15%)
Berlin [multiple stations] [*Jänicke et al.*, 2016] WRF 2012/07/01 - 2012/08/01	Noah LSM SLUCM BEP	0.5	0.9 -0.1 -0.6	0.4 (80%) -0.6 (-120%) -1.1 (-220%)	1.8	1.5 0.6 1.7	-0.3 (-17%) -1.2 (-67%) -0.1 (-6%)
Paris [Montsouris] [*Wouters et al.*, 2013] ARPS 2006/06/08 - 2006/06/13	LAICa	1.6	1.8	0.2 (13%)	4.3	5.1	0.8 (19%)
London [South Farnborough] [*Bohnenstengel et al.*, 2011] Met Office Unified Model 2008/05/07 - 2008/05/08	MORUSES	0.2	-0.6	-0.8 (-400%)	2.1	1.4	0.7 (-33%)

Table S4. Performance overview of different urban land-surface schemes coupled to convection-permitting models (CPM) in terms of day- and night-time urban heat island (UHI) magnitudes of mid-latitude European cities in the current and previous model evaluations. Units of the numbers are °C. The numbers in brackets are the percentual deviations between modelled and observed UHI magnitudes. This table is not a complete overview of all UHI studies or an intercomparison between the different models. Instead, it provides a benchmark of the capability of the CPM to represent urban heat islands. It should be noted that the differences between modelled and observed heat islands have several reasons, namely the difference in spatial representativeness between the model and the observations, the errors in model parameters, the approximations in the urban and atmospheric physics formulations, and/or errors in the observations.